Attorney Docket No.: 002-7: CSN 5387

PATENT APPLICATION

APPARATUS AND METHOD FOR USING A 2-WIRE BUS TO DESKEW 4 XAUI LANES ACROSS MULTIPLE ASIC CHIPS

Inventors: YU DENG, a citizen of CHINA, residing at 1046 Sandalwood Lane, Milpitas, California 95035

Assignee: CISCO TECHNOLOGY, INC. (A California Corporation)

Entity: LARGE

BACKGROUND OF THE INVENTION

[01] The XAUI (Ten Gigabit Attachment Unit Interface) is a 10-Gbps attachment interface unit for delivering 10-Gbps Ethernet speeds for chip-to-chip, board to board, and interbox communications. The XAUI spec defines four 3.125-Gbps streams for a total of 12.5-Gbps, which, taking into account 8B/10B encoding overhead, supports a 10 gigabit throughput with a maximum skew of 40 UI (unit interval).

[02] In a typical environment, 10- Gbps parallel data received on a wide parallel bus is serialized, 8b/10b encoded, and transmitted on the four 3.125- Gbps XAUI lanes. At the receiver the serial data is reformatted in parallel form for transmission on a parallel bus. If the routing of the four XAUI lanes is not closely matched then the data received on one lane will be skewed relative to data received on the other lanes. The XAUI spec provides for automatic deskewing of the lanes to eliminate the requirement of closely matching the routes for each lane. Skew is introduced between lanes by both active and passive elements of a XAUI link. The IEEE 802.3ae PCS deskew function compensates for all lane-to-lane skew observed at the receiver.

[03] XAUI is a self-timed interface having the timing clock embedded within the data. The data stream includes an alignment character (/A/) that is detected by a synchronization unit and used to align the data on the different lanes to deskew the data. Typically, data from each XAUI lane is buffered by a FIFO. The repetition of the alignment character (/A/) on each serial channel allows the FIFOs to remove or add the required phase delay to align the /A/ on each lane thereby deskewing the data on each of the four XAUI lanes.

[04] This alignment process is schematically depicted in Figs. 1A and B. In Fig. 1A the received data on each of the four XAUI lanes are skewed relative to each other. In Fig. 2B the data has been realigned to that the /A/ character in each lane is in the same column.

[05] For point to point connections, all the 4 serial lanes will end in the same chip, so that only a single PCS (Physical Coding Layer) deskew state machine, as defined in IEEE P802.3ae Figure 48-8, is required to deskew the 4 XAUI lanes. This specification is hereby incorporated by reference for all purposes.

[06] Implementation of 10Gbps Ethernet requires NxN switch fabrics capable of switching 10Gbps data streams. However, the implementation of an NxN 10GE switch fabric (N >= 64) as a single chip ASIC is either not feasible or just too expensive with present semiconductor fabrication methods because of the large number of ports. One solution is to implement the NxN 10 GE switch fabric as 4 NxN 2.5G chips with each chip operating on a single XAUI lane thus reducing the clock data rate on each chip by a factor of four. However, this creates the problem of deskewing 4 XAUI lanes across 4 different ASIC chips.

BRIEF SUMMARY OF THE INVENTION

[07] In one embodiment of the invention, a 2-wire bus links each of the NxN chips receiving a single XAUI lane to allow the deskewing of the data in the different chips. One of the chips is selected to be a master and asserts an alignment signal on a first wire of the bus when it detects an alignment character in the received serial data stream.

[08] In another embodiment of the invention, each of the slave chips detects when the alignment character is received in the serial data received on its respective lane. The time of detection is compared to with the time of assertion of the alignment signal by the master to determine the skew of data at each slave with respect to the data received at the master.

[09] In another embodiment of the invention, the internal clock speed of each NxN chip is twice the clock speed of 2-wire bus. The master generates an alignment signal that defines which phase of the 2-wire bus clock corresponds to the receipt of the alignment character.

[10] In another embodiment of the invention, each slave asserts an error signal on a second wire of the 2-wire bus if it detects that it has received the alignment character on a different clock cycle than the master.

[11] Other features and advantages of the invention will be apparent in view of following detailed description and appended drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[12] Figs. 1A and 1B are diagrams depicting the alignment of received data on different lanes ;

[13] Fig. 2 is a block diagram of 4 NxN switch fabric chips connected to implement an NxN switch fabric; and

[14] Fig. 3 is a block diagram of an embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[15] The invention will now be described with reference to specific embodiments by way of example not limitation. In the drawings like or similar parts in different views have the same reference number. In the following an embodiment will be described which is utilized in an NxN switch fabric. However, it will be apparent that the invention has general utility in many other environments.

[16] Fig. 2 depicts an NxN 10- Gbps switch fabric 10 implemented by four 2.5- Gbps NxN chips 12a-12d. As described above, each 10Gbps channel is serialized and transmitted over 4 XAUI lanes. Looking at a first channel, the first input XAUI lane, IX$_0$, is coupled to the first input port, I$_0$, of the first switch fabric chip 12a. The second, third, and fourth input XAUI lanes, IX$_1$, IX$_2$, and IX$_3$, are coupled to the first input port, I$_0$, of the second, third, and fourth switch fabric chips 12b, c, and d respectively.

[17] In general, each input channel has one of its four XAUI lanes coupled to a like-numbered port on each chip. Thus, the first input channel is coupled to input port I$_0$ on each of the four chips, the second input channel is coupled to input port I$_1$ one each of the four chips, and so on.

[18] Similarly, looking at a first output channel, the first output XAUI lane, OX$_0$, is coupled to the first output port, O$_0$, of the first switch fabric chip 12a. The second, third, and fourth output XAUI lanes, OX$_1$, OX$_2$, and OX$_3$, of the first output channel are coupled to the first output port, O$_0$, of the second, third, and fourth switch fabric chips 12b, c, and d respectively.

[19] Each switch fabric chip 12 can connect any input port to any output port. In the case of a 4-chip configuration, the controller switches the chips in tandem so that all 4 input XAUI lanes of a single input channel are switched to a single output channel. Thus, if the first input channel were to be switched to the first output channel port then port I$_0$ would be coupled to port O$_0$ on each switch fabric chip. In general, the four input XAUI lanes of any input channel can be coupled to the four output XAUI lanes of any output channel. Additionally, each port can be coupled to an input and output XAUI lane to provide full duplex switching.

[20] As described above, with reference to Figs. 1A and B, the data transmitted in the XAUI lanes may be skewed. When the XAUI lanes are received in a single chip the standard deskew state machine, as defined in IEEE P802.3ae Figure 48-8, is utilized to deskew the 4 XAUI lanes. However, in the present embodiment, the lanes are received at different chips so that a new method of deskewing needs to be implemented.

[21] In the presently defined embodiment, a complete XAUI deskew state machine is included for each port on each of the four switch fabric chips. However, since each of the XAUI lanes of a single channel are received on a different chip there it is impossible for the deskew state machines to synchronize the data in the different XAUI lanes because the /A/ character in the different lanes are detected on different chips. A technique for allowing the XAUI deskew state machines on different chips to synchronize data on different XAUI lanes received on the different chips will now be described.

[22] In the currently described embodiment, a 2 wire bus, implemented as traces on a printed circuit board (PCB), is used to communicate deskewing information between the separate chips coupled to the XAUI lanes. The 2 wires of the bus are: Align_Char (A_C): Bi-directional I/O with Output_Enable, and Error_Ind (E_I): Bi-Directional I/O with Output_Enable.

[23] Fig. 3 is a block diagram of the functional units included in each of the switch fabric chips 12. Each chip includes a clock generator that receives a 156.25Mhz external reference clock. The traces to each chip are of the same length the external clock signals received at each chip are in phase. The clock generator includes a phase-locked loop (PLL) 30 that generates an in-phase bus clock signal (clock156 = 156.25Mhz) and an internal chip clock signal (clock312 = 312.5Mhz). The bus clock runs at a slower speed than the chip clock due to the physical characteristics of the PCB.

[24] The incoming data is compared to the /A/ character in a detector block 32 including a first comparator 36 and a first flip-flop 38 having its output connected to the trigger input of a bus driver control unit 40. The first flip-flop is clocked by the internal chip clock and asserts an /A/ detection signal at its output when /A/ is detected in the received data stream. The output of the bus driver control unit is coupled to the A_C bus wire via an output driver 42 including an output enable (O/E) input.

[25] A phase detection block includes a second flip-flop 44, clocked by the internal chip clock and having a delayed bus clock as its input, and a first AND gate 46. The outputs of the first and second flip-flops 38 and 44 are connected to the first AND gate 46 which has a phase indicating signal as its output. The phase indicating signal is coupled to phase input of the bus driver control unit 40.

[26] Each chip also has an E_I signal generator that includes an A_C decoder 50 having a signal input coupled to the A_C bus wire via a third flip-flop 52, clocked

by the internal bus, a clock input receiving the bus clock, and an output for generating a timing signal. The E_I signal generator also includes a second comparator 54 having a first input coupled to receive the on chip /A/ character signal detection signal with exactly via matched pipe-line delay element 55, for compensating the delays introduced into the received A_C signal, and a second input coupled to receive the timing signal from the decoder. The output of the second comparator is coupled to the E_I wire of the two wire bus via a fourth flip-flop 56 and an output driver 58. The output of flip-flop 56 is connected to both the input and output/enable of the output driver 58, so that it will drive E_I active only when error is detected, otherwise it will tri-state E_I.

[27] As described above, each chip includes a deskew state machine 60coupled to receive A_C and the E_I signal from the two-wire bus.

[28] The operation of the system will now be described. Turning first to the generation of the A_C signal. One of the chips, for example the first chip 12a, functions as a master and asserts the A_C signal on the A_C wire of the two wire bus when it detects the /A/ character in its received data stream. The O/E input of the A_C bus driver 42 is set high during configuration. The remaining chips have their O/E disabled and function as slaves. The slaves adjust their data streams so that the /A/ character is aligned at each chip (as depicted in Fig. 1).

[29] The A_C signal is driven high, on the master chip 12a, when the first comparator 36 asserts its output signal indicating that the /A/ character has been detected in the received data stream. This output is synchronized to the internal clock by the first flip-flop 38. Thus the output of the first-flip 38 occurs during a detection internal clock cycle that indicates when the /A/ character was detected on the master chip 12a.

[30] Since the bus is sampled only once every two internal chip cycles, information regarding the phase of the bus clock corresponding to the detection internal chip cycle is encoded onto the A_C bus line. In this embodiment, the second flip-flop 44 outputs the phase of the bus clock for each chip clock cycle. If /A/ is detected when the phase of the bus clock signal is high then the output of the first AND gate (the phase signal) is high at the time of detection and if /A/ is detected when the phase of the bus clock signal is low then the AND gate output is low at the time of detection. The bus driver control unit drives asserts A_C for one bus clock cycle if /A/ is detected when the bus clock phase is low or for two bus clock cycles if /A/ is detected when the bus clock phase is high.

[31] The detection of the A_C signal at a slave chip will now be described. The A_C bus wire is coupled to the A_C decoder 50 by the third flip-flop 52 that synchronizes the A_C signal to the internal chip clock. The A_C decoder 50 generates an A_C synch signal corresponding to the phase of the bus clock signal encoded onto the A_C bus signal.

[32] The generation of the E_I signal at a slave chip will now be described. The second comparator 54 generates an error indicating signal if the generated A_C synch signal and on-chip generated /A/ detection signal are mismatched. This error indicating signal is input to the E_I bus driver to drive the E_I bus high when error is detected, otherwise it will tri-state its E_I bus driver. The Error_Ind bus has a pull-down resistor connect to GND on PCB. Thus, multiple chips can drive the bus if mismatches are detected.

[33] Thus, each slave chip 12b to 12d that detects a mismatch will drive its Error_Ind pin active for one bus clock cycle. For slave chips with no mismatch, the Error_Ind pin is tri-stated. The PCS deskew state machines (IEEE P802.3ae Figure 48-8) 60 in all 4 chips will use the same Error_Ind input as their "deskew_error" input, by doing it this way, the deskew state machine in all 4 chips will be always in the same state. And Finally, all the 4 chips will align with lane 0.

[34] Each chip supplies its decoded A_C signal and received E_I signal to its deskew state machine 60 to provide the timing required for the deskew state machines to align the data streams on all chips four chips 12a-d.

[35] This embodiment of the invention has the advantages of using the smallest number of pins for inter-chip connections, reducing the chip package size (cost and PCB real estate), and facilitating easier PCB routing. Additionally, the system is simple and standard compatible. Only a single IEEE P802.3ae Figure 48-8 PCS deskew state machine is needed for each port. Further, the system is symmetric. In can be utilized with an NxN 10GE switch fabric with 4 chips, an N/2 x N/2 10GE switch fabric with 2 chips, or an N/4 x N/4 10GE switch fabric with 1 chip by using exactly the same deskew scheme.

[36] The invention has now been described with reference to the preferred embodiments. Alternatives and substitutions will now be apparent to persons of ordinary skill in the art. For example, the particular synchronization and logic elements are exemplary and various other substitute techniques known in the art can be utilized. Accordingly, it is not intended to limit the invention except as provided by the appended claims.